Energy Reports 7 (2021) 3775–3784

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

Classified atmospheric states as operating scenarios in probabilistic power flow analysis for networks with high levels of wind power

Amaris Dalton [a,*], Bernard Bekker [a], Matti Juhani Koivisto [b]

[a] Department of Electrical and Electronic Engineering, Stellenbosch University, South Africa
[b] Wind Energy Department, Technical University of Denmark, Risø Campus, Denmark

ARTICLE INFO

Article history:
Received 23 October 2020
Received in revised form 26 March 2021
Accepted 14 June 2021
Available online xxxx

Keywords:
Probabilistic power flow analysis
System simulation
Wind power
Meteorology
Atmospheric states

ABSTRACT

Large-scale atmospheric circulation patterns are the primary drivers of wind power variability on power networks at timescales of hours to days. This paper proposes a methodology that allows power system operators and planners working on networks with high levels of wind generation, to conduct probabilistic power flow (PPF) analyses by defining network 'operating scenarios' – i.e. the probability density functions of generators, and correlations between generators representative of a future system state – based on concurrent classified atmospheric states. The most significant contribution made by this paper is in illustrating how PPF operating scenarios derived from clustering historic generation data as a function of a set of classified atmospheric states reduces simulation uncertainty within a PPF analysis. It is anticipated that the proposed methodology may provide network planners with more appropriate operating scenarios for PPF analyses when compared to an unclustered base state, and may assist network operators in converting wind power point-forecasts into probabilistic forecasts whereby the spatial correlations between generators are incorporated. This methodology is illustrated through a case study considering 11 geographically disperse wind generators on the South African transmission network.

1. Introduction

1.1. Power flow analysis and network uncertainty

Power flow analysis is an important tool for determining the static operational conditions within various nodes of an interconnected electricity network. A power flow analysis is conducted by numerically solving a set of non-linear algebraic power flow equations (Stott, 1974) to determine the value of relevant network parameters, including voltage magnitude, phase angle, and active- and reactive power. On shorter time-scales (i.e. day-to-day basis) network operators have used power flow analysis in conducting contingency analysis (Van Hertem et al., 2005), determining generation scheduling and unit commitment requirements (Raglend and Padhy, 2006), network congestion and loading relief (Min et al., 2008). At longer time-scales, power flow analysis is useful to network planners in determining grid strengthening requirements and the layout of future network topologies (Bent et al., 2012).

Stochastic behavior of both consumer load demand and generation components within a power system results in a negative economic impact, largely due to the implicit difficulties in balancing the system (Ueckerdt et al., 2015). Uncertainty in power networks may be resolved by way of either a deterministic power flow (DPF) or probabilistic power flow (PPF) analyses. In traditional power systems, the consumer load had been the primary stochastic component within the system, as thermal generators behave deterministically in normal operation. Thereby, a DPF would be conducted for a set of system 'snapshots' (Jiang et al., 2013) or operating scenarios to capture network conditions resulting from typical variations in load demand such as weekday load peak versus weekend load peak, or summer versus winter. Two important drawbacks of DPF to note are that: uncertainties are not adequately taken into account as the probabilistic nature of renewable energy sources remain unrecognized (Tang et al., 2016); and system snapshots are defined arbitrarily, based on the experience of the system engineers (Borkowska, 1974).

Increasing levels of variable renewable energy (VRE) on power networks result in increased variability in the power flows in the network (Kroposki et al., 2017). Indeed, in networks with high levels of VRE penetration, VRE has become a larger source of uncertainty than load, as VRE resources are inherently more variable. The combined effect of VRE and load stochasticity results in a very large set of possible power flow scenarios. Thereby, traditional DLF operating scenarios are becoming insufficient in representing uncertainties within power systems (Chen et al., 2008).

* Corresponding author.
E-mail addresses: amaris_dalton@sun.ac.za (A. Dalton), bbekker@sun.ac.za (B. Bekker), mkoi@dtu.dk (M.J. Koivisto).

https://doi.org/10.1016/j.egyr.2021.06.060
2352-4847/

Additional reasons for DLF insufficiency include the complexity and computational burden resultant from large multivariate problems, and the aggregated uncertainty resulting from the deterministically uncharacterizable dependencies between system load, solar PV and distributed wind power (Haghi et al., 2010). As an alternative to DLF, PPF is becoming increasingly popular amongst power system planners and operators (Chen et al., 2008). PPF accounts for network uncertainties through the introduction of probability calculus whereby relevant variable network parameters such as load and generation are represented as distribution functions rather than single point-values. The points on such distribution functions are then solved in what is most often a stochastic Monte-Carlo type iterative calculation (Usaola, 2009). The outputs from such calculations are thus also probabilistic and provide a representation of the uncertainties associated with these outputs.

1.2. PPF challenges and intended applications

This paper proposes a novel methodology for deriving power system operational states based on atmospheric circulation. These operational states are represented as the joint probability density functions (PDF) (or copula functions) of multiple geographically dispersed wind generators along with the correlations between these generators. Stochastic variables such as wind power can effectively be represented as PDFs, as such high-frequency uncertainties correspond to recurrent conditions (Fitiwi et al., 2015). In terms of applications, the proposed methodology distinguishes between long term (weeks to months) 'planning' PPF and short term (hours to days) 'operational' PPF. Wind power variability at time-scales of hours to days are associated with the passage of large-scale weather systems (Dalton et al., 2019; Kiviluoma et al., 2016; Zhang et al., 2014). The primary power system applications to be informed by an operational PPF at such time-scales include day-ahead electricity market, reserve requirement, unit commitment, economic dispatch, and short-term maintenance scheduling (Zhang et al., 2014). Large-scale weather systems may also provide a useful perspective from which to understand network variability at longer timescales, thereby informing network planning studies (Grams et al., 2018). At timescales of weeks to years, the primary power system applications that may be informed by a network-planning PPF include network strengthening requirements, expansion planning, and generator siting. It is anticipated that the proposed methodology would have applications in both operational PPF (time-scales of hours to days) and planning PPF (time-scales of weeks to years) as explained below.

1.2.1. Operational probabilistic power flow analysis

If conducting PPF analyses at operational time-scales of hours to days, deriving a copula function for multiple wind generators essentially becomes a probabilistic forecasting problem. A reasonable point of departure may be to consider statistical regularities in historic wind generation data arising from diurnal and seasonal cycles. Indeed, at the time-scales of 30 min to 6 h, it has been shown that most wind power forecasts tend to be stochastic (Zhang et al., 2014), making the use of either statistical approaches (e.g. autoregressive techniques) or artificial neural networks (e.g. feed-forward, ADALINE, etc.) feasible (Soman et al., 2010). At timescales of more than 6 h however, dynamic physical weather models tend to be preferred to statistical approaches. As this study is focused on the role of synoptic-scale atmospheric circulation in the electricity network, such dynamic weather forecasts would have to be incorporated into the analysis.

Wind power forecasts derived from converting wind speed forecasts to wind power are often represented as a deterministic single value point or a 'point-forecast' representing a specific look-ahead time. In a liberalized electricity network with multiple wind power plants operated by private companies, wind power forecasts are typically submitted by these companies to the central system operator. From the perspective of the system operator, a couple of problems may arise when using point-forecasts in operational decision making as discussed by Botterud et al. (2011). Pertaining specifically to PPF, a point-forecast does not provide a representation of forecast uncertainty, as needed in a PPF, unless a computationally expensive ensemble forecast is generated. Secondly, point-forecast skill will differ between locations based on the forecast models used, local terrain complexity and the quality of local observations. Furthermore, in conducting a PPF for a network with high levels of correlated wind power, it is necessary for the data interdependency structures (i.e. the level of correlation) between geographically dispersed wind generators to be taken into account. The level of correlation between wind generators informs the smoothness of the aggregated power profile fed into the network (Monforti et al., 2014). Wind power forecasts are however likely to be provided to the grid operator by independent parties responsible for operating wind generators on a network (Zhang et al., 2014). Therefore it is unlikely that the spatial correlation between several geographically dispersed generators will be accurately represented for a future system state.

Within the context of operational PPFs, which are informed by wind power forecasts as described above, the problem that this paper seeks to address may be summarized as: firstly how to generate a probabilistic wind power forecast for a network that contains multiple geographically dispersed wind generators, where the forecast contains a valid assumption as to the shape of the PDF and the correlations between wind farms using a NWP point-forecast; and secondly illustrating of such a forecast may be incorporated as an operating scenario (i.e. a copula function representing the future system state) into an operational PPF.

1.2.2. Planning probabilistic power flow analysis

System reliability/adequacy indices are important outcomes of PPF analyses in network planning studies (Chen et al., 2008). These indices may provide the planner with information such as the probability, frequency and duration of periods during which system reliability will be at risk. Traditionally, network planning is based on the 'worst-case-scenario principle' used to meet network requirements throughout most circumstances (Repo and Laaksonen, 2005). The worst-case scenarios are typically very conservative estimations of the maximum load combined with minimum production or minimum load combined with maximum production. These assumptions are likely to be invalid for networks with large amounts of VRE generation where the stochastic properties of generators need to be taken into account (Repo and Laaksonen, 2005).

However, in system planning studies only a limited number of network operating scenarios can be simulated as part of an analysis, particularly in large networks (Leite da Silva et al., 1990). It is thereby important that an appropriate selection be made of the most relevant network operating scenarios to be used in the PPF. In networks with high levels of wind power penetration – i.e. whereby atmospheric circulation is a dominant instigator of power variability – it is anticipated that classified atmospheric states may be used by network planners in deriving appropriate operating scenarios to be included in the analysis. The proposed methodology may inform appropriate operating scenario selection by providing planners with a probabilistic valuation of atmospheric states most representative of regional climatology based historic occurrence frequency, and in terms of the identification of atmospheric states that the network may find particularly challenging to accommodate (e.g. strongly correlated wind conditions).

An important distinction between the operational and planning PPF applications of the proposed methodology is that the operational PPF application is reliant on NWP forecasts (to determine a future system state) and historic wind power data (to determine probabilistic properties), whereas the planning PPF application only requires historic wind power generation data. In instances where measured wind power data is not available due to e.g. confidentiality reasons, or when a network expansion is planned, the simulation of long-term generation data becomes necessary.

1.3. Paper overview

The primary underlying hypothesis of this paper is that: in planning and operational PPF, the probabilistic spatial–temporal information needed to select appropriate operating scenarios reflective of future system states so as to resolve the aforementioned problems, may be obtained through the clustering of historic wind power time-series for multiple generators as a function of concurrent classified atmospheric states. This paper builds on work present by the authors in (Dalton et al., 2020a)

To test this hypothesis, a methodology is proposed that consists of the following steps: (a) classification of NWP reanalysis data using Self Organizing Maps to represent a set of atmospheric states reflective of regional climatology; (b) simulation of a wind power time-series for multiple geographically dispersed generators for a period and time-step concurrent to that of the NWP reanalysis dataset (to be implemented in planning assessments where the wind generators do not exist yet or operational instances when measured data is not available); (c) clustering wind power times series (simulated or measured) as a function of the classified atmospheric states and using the resultant clusters to derive a set of copula functions which are in turn used as inputs into a PPF. An overview of the paper outline is provided in Fig. 1.

The layout for the remainder of this paper is as follows: Section 2 provides a theoretic overview of the elements used within the proposed methodology, Section 3 introduces a case study of 11 wind farms in South Africa whereby the methodology is illustrated, Section 4 shares results and discussions and Section 5 concludes the paper.

The most significant contribution made by this paper is in illustrating how PPF operating scenarios –represented as copula functions – can be derived from clustering historic generation data as a function of a set of classified atmospheric states so as to reduce uncertainty within the PPF analysis. These classified atmospheric states represent the large-scale atmospheric circulation of an area, which allows for the inputs into PPF to be based on physical meteorological phenomena. It is anticipated that defining network states in terms of atmospheric circulation may provide network planners with a method for selecting appropriate operating scenarios to include in a PPF analysis, and may assist network operators in converting wind power point-forecasts into probabilistic forecasts whereby the spatial correlations between generators are incorporated.

2. Methodology

2.1. Scenario selection and self organizing maps

Classification of atmospheric states is done using Self-Organizing Maps (SOMs). SOMs are a type of artificial neural network that learns in an unsupervised environment (Kohonen et al., 1990). SOMs provide a method for clustering high dimensional data into (what is typically) a 2-D node lattice, wherein the topological features of the input data (or feature vectors) are maintained. The first step within the SOM training process consists of defining a set of nodes (or reference vectors) of equal dimension to the feature vectors. Competitive learning is then utilized whereby the Euclidean distance between a feature vector and the defined set of reference vector(s) is measured. The winning node, or *Best Matching Unit* (BMU), is the node whose weight vector is closest to that of the feature vector. Once the BMU has been identified, its weight along with the weights of the nodes in its topological neighborhood on the SOM grid, is updated towards that of the feature vector. SOM training is an iterative process and is completed once changes to node locations no longer are being made. Mathematically the update of a neuron n may be represented as:

$$V_n(t+1) = V_n(t) + \varphi(i,j,t).\beta(t).(D(s) - V_n(t)) \quad (1)$$

where: V_n is the weight vector; t is the step-index (or current iteration); φ is the neighborhood function providing the time-varying form of topological neighborhood from the BMU (neuron i) to neuron j for the t'th iteration; β is the temporally decreasing learning function; and $D(s)$ is the randomly selected feature vector (Kohonen and Honkela, 2007).

SOMs were selected as a clustering methodology for atmospheric data in favor of other commonly employed clustering methods such as K-means or principal component analysis (PCA), as SOMs do not discretize the data through operations such as Eigenvector analysis or correlation, nor does it force orthogonality (Lennard and Hegerl, 2014; Reusch et al., 2005). Rather, SOMs consider the atmospheric input data as a continuum that is capable of recognizing non-linear relationships within the data. The end result of the SOM may be interpreted as a generalized set of smoothly transitioning atmospheric states, rather than patterns of variance, as would be the case with PCA (Reusch et al., 2005). Accordingly, SOMs have been broadly applied within the atmospheric sciences as discussed by Liu and Weisberg (2011) and Sheridan and Lee (2011).

In the final step of the SOM classification, the reanalysis time-series is clustered in terms of the set of classified atmospheric states – i.e. a SOM node number is assigned to each time-step. The process for doing this is similar to the SOM training process whereby the weighted Euclidean distance between each time-step and each classified SOM node is calculated, and the BMU is accordingly identified. Each time-step is then assigned a SOM node label. When this methodology is implemented operationally, NWP forecasts will similarly be used to determine future atmospheric states as per the SOM classification. The similarity between neighboring SOM nodes is anticipated to give the proposed methodology some robustness against NWP forecast errors whereby a NWP forecast with a certain forecast error will still be attributed to a SOM node within the correct neighborhood. This is thereby deemed to fulfill a similar function to a probabilistic ensemble forecast in determining the predictive distribution.

2.2. Simulation of wind power

For operational PPF, it is anticipated that the simulation of wind power time-series would be largely unnecessary, as the grid operator would have access to measured wind power data. For research purposes, such data are usually proprietary and therefore unavailable. Similarly, for network expansion planning PPF studies, wind measurements may not be available for sites considered for wind farm siting (Al-yahyai et al., 2010). This lack of measured data necessitates a reliable and long term wind power simulation to represent variability in the network. Accordingly, a simulation of wind power time-series for a number of wind power generators is included in this methodology. The simulation time-series

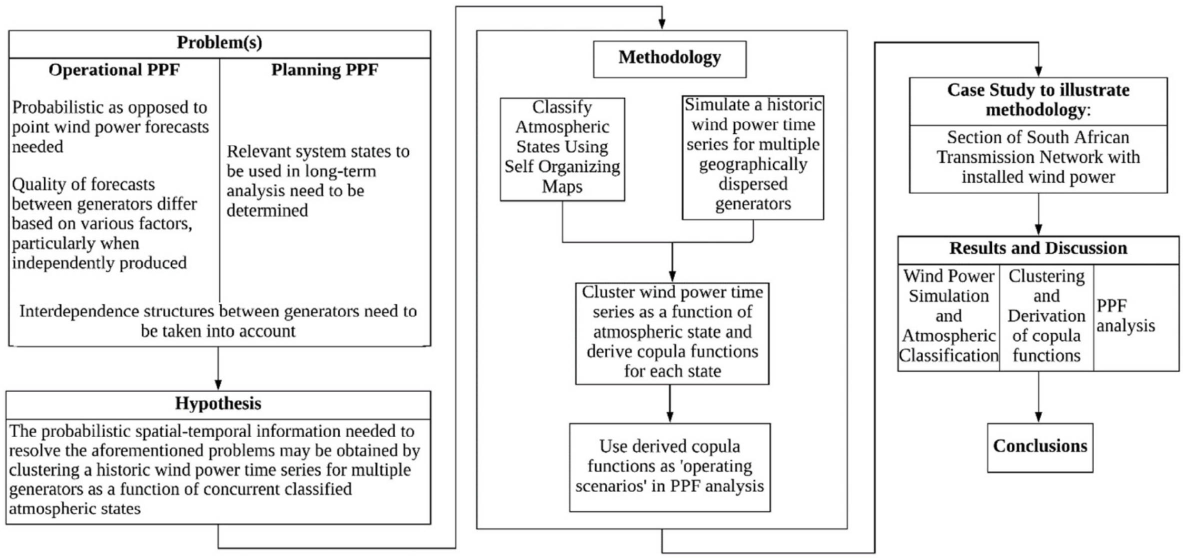

Fig. 1. Paper outline.

is concurrent in terms of period and time-step to the reanalysis time-series used in the previous step of the methodology. The period and time-steps will be case study dependent (consider Section 3). A brief description of the simulation methodology is outlined below.

Wind power time-series are simulated using the CorWind model developed by DTU as part of the CorRES (Correlations in Renewable Energy Sources) framework. To rectify the mesoscale smoothing effect associated with NWP modeling (Larsén et al., 2012) which results in a spectral energy deficit, CorWind adds an element of stochastic variability to modeled numeric wind data. The combination of the meteorological data stochastic simulation is presented (Koivisto et al., 2020), with validation for Denmark. These fluctuations are informed by power spectral densities, the details of which may be seen in Sørensen et al. (2008, 2002). The CorWind model was validated also for South African conditions in (Dalton et al., 2020b) and has been used in various energy modeling studies e.g. Koivisto et al. (2019) and Sørensen et al. (2018).

CorWind allows for the conversion of wind speed to wind power through the specification of a power curve. As this study considers wind farms rather than single turbines, standard power curves provided by turbine manufacturers were transformed into multi-turbine power curves using the method by Norgaard et al. (2004). The multi-turbine power-curve effectively smooths out certain short-term fluctuations to provide a wind power profile that is representative of an aggregation of multiple generators, as would be the case with a wind farm.

Once the simulations have been completed, the SOM node labels assigned to each time-step within the reanalysis dataset is carried over to the corresponding simulation time-series. Accordingly, the wind power simulation time-series is clustered according to a corresponding set of atmospheric states.

2.3. Copula functions

In a review article on probabilistic wind power forecasting techniques by Zhang et al. (2014), it was noted that a spatiotemporal forecasting process may be most suitable when dealing with multivariant PDFs. A copula is a cumulative distribution function that takes marginal – or unconditional - distributions and returns a single multivariate distribution that contains the dependence structure between the individual margins. The margins in turn are the inverse normal CDF (if it exists) - which is used to transform the uniform distribution back to its actual domain. When considering VRE, copula modeling is advantageous as it separates the individual marginal distribution of generators from the dependencies between various locations (Koivisto et al., 2016).

In terms of finding a marginal distribution that is suitable for wind power modeling, it should be noted that parametric distributions such as Gaussian, Beta or Weibull do not provide a good fit. Therefore the non-parametric empirical CDF (ECDF) marginal was fitted to the wind power data. The ECDF is a step-function that is essentially identical to the CDF of the variable.

For a d dimensional continuous random variable X where $d \geq 1$ *and* $X = [X_1,\ldots, X_d]$, the CDF representing the probability that $X_d \leq x_d$ may be expressed as:

$$F(x) = P(X_1 \leq x_1, \ldots, X_d \leq x_d) \tag{2}$$

Accordingly, Sklar's theorem (Sklar, 1973) states that there exists a copula function C, whereby the d-dimensional joint distribution H, with 1-margins $F_1,\ldots, F_d$, can be written as:

$$\begin{aligned} H(x_1, \ldots, x_d) &= P[X_1 \leq x_1, \ldots, X_d \leq x_d] \\ &= C[F_1(x_1), \ldots, F_d(x_d)] \end{aligned} \tag{3}$$

Thereby the C copula functions provide the dependence structure between components and the marginals providing the CDFs of $X_1,\ldots, X_d$.

There are several families of copulas, including the Archimedean, Elliptical and Gaussian copulas. It has been observed in previous studies that variations based on the selection of copula type in modeling wind power distributions are rather insignificant (Leuthold et al., 2008). Gaussian copulas were observed to be the main approach for characterizing the interdependence structure of multiple wind farms (Zhang et al., 2014) and was therefore derived for each SOM cluster within the wind power simulation time-series. A Gaussian copula with correlation matrix P may be expressed as:

$$C_P^{Gauss}(u) = \varphi_n(\varphi^{-1}(u_1), \ldots, \varphi^{-1}(u_d); P) \tag{4}$$

where φ_P is a multivariate normal CDF over P^d with expectation mean vector zero, and φ^{-1} is the inverse CDF for the standard Gaussian distribution

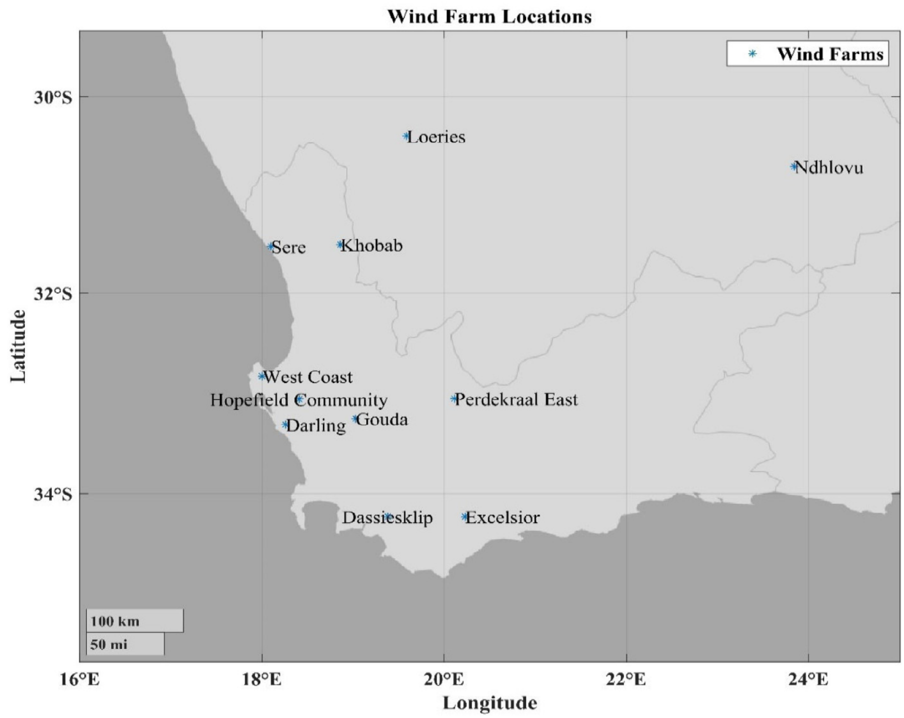

Fig. 2. Location of geographically dispersed wind farms used in the case study.

3. Case study: PPF analysis of the western cape transmission network, South Africa

Hourly wind power data was simulated for 11 wind farms located within the Western and Northern Cape provinces of South Africa for a period of 5 years (2010–2014). As is evident from Fig. 2, the wind farms are dispersed across a fairly large geographic area. To avoid confusion it is important to note that for the simulation period most of these wind farms were not as yet operational — rather the 5 years simulation data is used to derive the probabilistic properties of these generators, which are still deemed to be representative of current conditions. The simulation period was selected based on the reanalysis data period within CorWind for South Africa.

The 850 hPa geopotential heights from the open-source ERA5 (Copernicus Climate Change Service, 2017) reanalysis dataset were used as a classification parameter for the SOMs. The geopotential height data was downloaded for the same period (2010–2014) and time-step (hourly) as the wind power simulation data. Geopotential height refers to the elevation above sea level where the specified pressure level is found. 850 hPa Geopotential heights were selected as the input parameter to the classifications schema as it provides a good representation of synoptic-scale circulation patterns in that it reflects well the baroclinic and barotropic systems which are features of South Africa's climatology (Tyson and Preston-Whyte, 2000). Furthermore, 850 hPa provides sufficient elevation to describe circulation above the complex local geography of the South African escarpment.

The selection of SOM size (i.e. the number of SOM nodes) is a subjective decision (Lennard and Hegerl, 2014) which depends on the intended application and the amount of detail or generalization required. Thereby the smaller the SOM, the greater the degree of generalization and the larger the SOM, the greater the level of detail provided. In the practical application of this methodology, this decision would ultimately lie with the end-user, e.g. the network operator. The decision would be based on a trade-off between desired operational complexity (i.e. number of operating scenarios) and accuracy of probabilistic representation (i.e. number of SOM nodes). SOM size selection would also be a function of regional climate variability whereby a greater SOM size might be required to resolve a complex thermally driven climate. To illustrate this methodology, a six by six SOM node topology (i.e. 36 nodes) was selected. A hexagonal lattice structure was selected in favor of a rectangular lattice so as to increase the level of interdependence between SOM nodes. It has however been shown that the selection of the lattice shape has little bearing on the final SOM (Openshaw, 1994). The classification area was bounded between 22–40°S and 5–40°E, making the SOM region significantly larger than the geographic spread of the wind turbines for which the simulation was conducted. This was done to allow for large-scale circulation associated with the Indian and Atlantic oceans to be included in the classification.

Gaussian copula functions were derived using the clustered wind power time-series, for each SOM node as discussed within the Methodology section. For the purposes of the PPF, DIgSILENT PowerFactory software was used. The derived Copula functions were uploaded into PowerFactory as 'characteristics', which were assigned to appropriate generators. Thereby each wind power generator had a set of distribution functions for each SOM node (or atmospheric state) and each SOM node had a corresponding correlation matrix for all generators.

For an accurate representation of the current operational transmission networks within the study area, South Africa's national network operator Eskom has supplied the authors with the appropriate PowerFactory case files. Due to the size of the South African network and due to the relatively low penetration of wind energy in the network (3.8% for 2019 BP, 2020), parts of the network were isolated so that the experiment being conducted would be representative of a network with a more significant share of wind energy. Accordingly, for the network configuration used in the case study, wind energy contributed approximately 10% to installed generation capacity. Only wind generators were considered probabilistically as variable inputs. Loads, along with the balance of the generators on the network, were considered

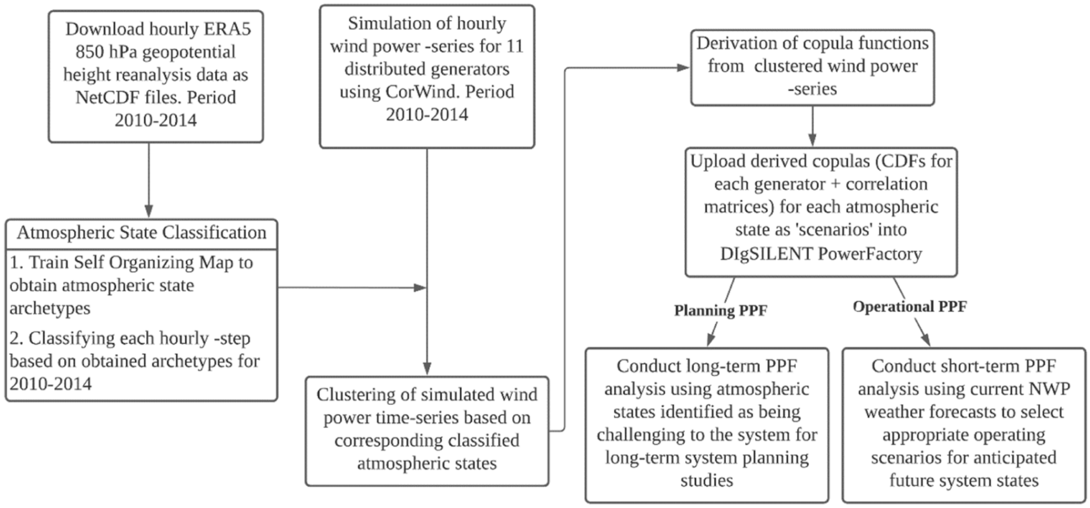

Fig. 3. Case study methodological outline.

deterministically as point values. The commonly used Newton–Raphson power flow method was employed using non-linear AC power flow equations. With the aim of reducing the computational burden, the quasi-Monte Carlo sampling method was used, which has been shown to lead to faster convergence (DIgSILENT GmbH, 2019). The idea behind the quasi-Monte Carlo method is that sampling space is more uniformly covered which avoids sample clustering. PPF analysis was conducted for 1000 sampling iterations for each of the atmospheric states.

The methodology as applied in the case study is summarized in Fig. 3 below.

4. Results and discussion

4.1. Classified atmospheric states

Fig. 4 shows the 36 (6 × 6) node SOM wherein the classified modes of the 850 hPa geopotential heights is used to represent characteristic circulation over the study area. Simply put, each SOM node represents a classified atmospheric state. The frequency of SOM node occurrence is indicated as a percentage above each node. As anticipated, it is evident that the level of dissimilarity between nodes increases with the distance between nodes. Nodes on the left of the SOM lattice are dominated by high pressure circulation whereby the upper left-hand corner represents the ridging of the semi-stationary Atlantic Ocean High Pressure System onto the sub-continent. The bottom left hand corner in turn represents the ridging of the Indian Ocean High Pressure System. Conversely, nodes towards the right of the SOM lattice are dominated by low pressure circulation, notably the bottom right-hand corner which represents the passage of mid-latitudinal cyclones. These mid-latitudinal cyclones and associated cold fronts have been shown to be significant instigators of strong wind conditions and wind power ramp events throughout the study region (Dalton et al., 2020b; Kruger et al., 2010).

4.2. Wind power variability for different atmospheric states

Fig. 5 shows correlation matrices for a subset of five sampled wind farms whereby their respective generation time series were clustered based on the concurrent occurrence of four atmospheric states as represented by SOM nodes (1,1), (1,6), (5,1) and (6,6). The four SOM nodes displayed were selected based on the frequency of occurrence and because of their relative location on opposite sides of the SOM map. Along the diagonals, variable histograms are displayed. Simply put, Fig. 5 shows how the probabilistic distribution of a generation time series, and the correlation between wind farms, differ based on the atmospheric state. From Fig. 5 it is thereby evident that the proposed methodology of clustering of wind power time-series based on atmospheric state, is able to elucidate significant differences in probabilistic properties of wind generators, and in the data interdependency structures between generators. It is evident that SOM node (1,1) is associated with weakly correlated generation between turbines along with a high probability of low power generation for each of the turbines sampled. Conversely SOM node (6,6) is associated with a high correlation between turbines and a high probability for high power generation for each of the turbines.

4.3. Probabilistic power flow analysis

Fig. 6 shows active power in MW on two sampled transmission lines for the PPF conducted, based on atmospheric states represented by SOM nodes (1,1), (1,6), (5,1) and (6,6), compared to a base scenario, a best-case scenario and worst-case scenario. The base state copula was derived without considering any clustering criteria, but rather considering the entire simulation period. The best-case and worst-case scenarios are reflective of the highly conservative, but commonly employed 'worst-case planning principal' as discussed by Repo and Laaksonen (2005), whereby the best-case scenario load flow analysis considers all wind power generators to be fully correlated and operating at full capacity. Conversely, the worst-case scenario considers all wind power generators to be generating no power. Note that due to non-disclosure restrictions, the line/component names provided in Figs. 6 and 7 only refer to the general geographic area within which the line/component is located.

For Stikland, the base state is associated with comparatively high network simulation uncertainty as is evident from the relatively flat PDF curve which assigns the same approximate probabilities to a relatively broad range of potential line loadings (approx. 100–120 MW). When considering the specific operating scenarios, simulated network uncertainty tends to be decreased as is evident from higher probabilities assigned to a relatively narrow range of potential line loadings. Indeed, strong cyclonic circulation (SOM (6,6)) is associated with a high probability of relatively low active power (approx. 85 MW) and ridging high pressure circulation (SOMs (5,1) and (1,1)) is associated

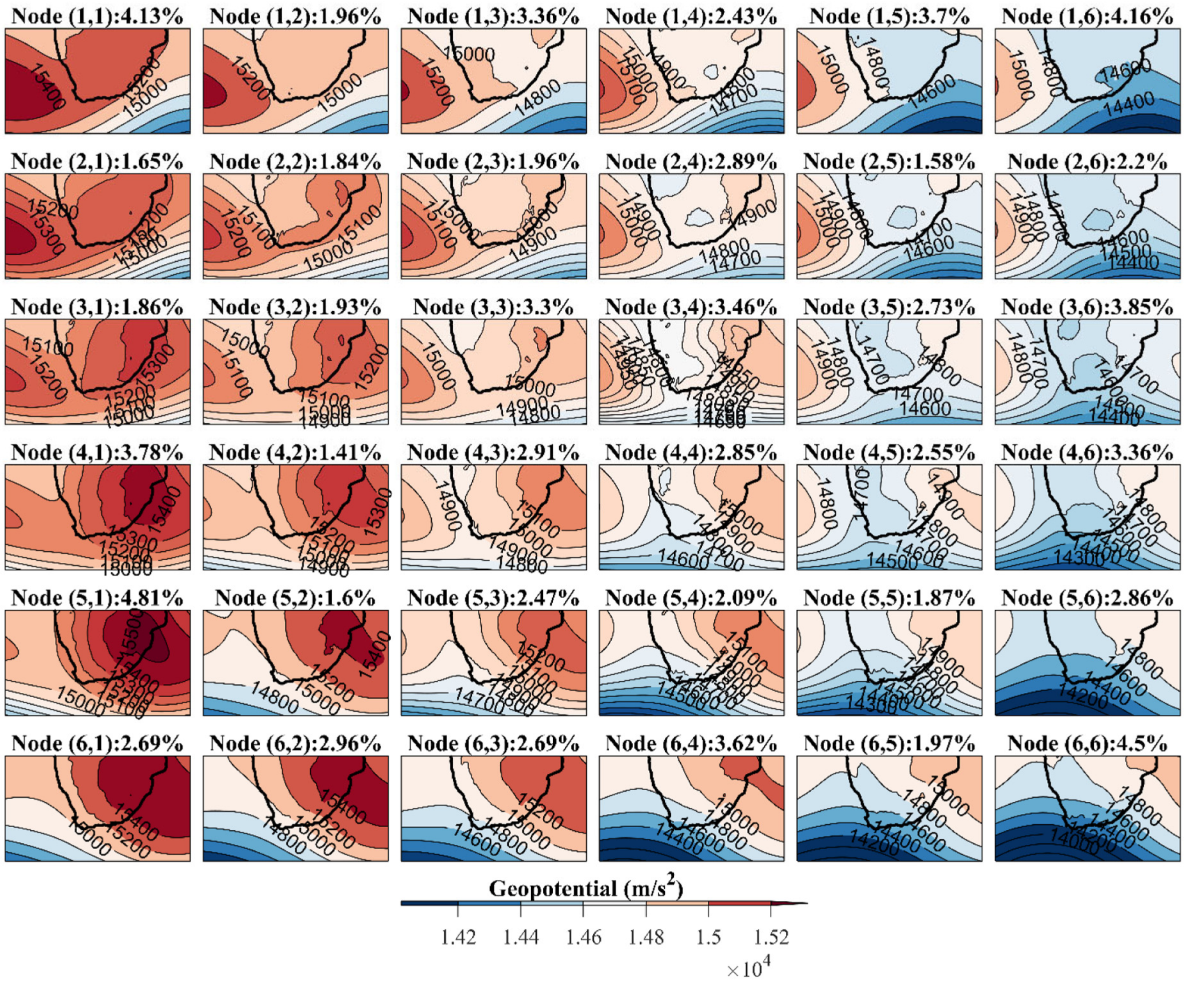

Fig. 4. SOM of 850 hPa geopotential heights, along with the frequency of SOM node occurrence over Southern Africa (2010–2014).

with a high probability of relatively high active power (approx. 115–125 MW).

For Acacia, it is similarly apparent that the operating scenarios based on atmospheric circulation reduce network uncertainty when compared to the base state. Indeed, strong cyclonic circulation (SOM (6,6)) is associated with a high probability of relatively high active power (approx. 165–170 MW) and high pressure conditions (SOMs (5,1) and (1,1)) are associated with a high probability of relatively low active power (approx. 135–140 MW). SOM (1,6) was largely unable to differentiate itself from the base state for both Stikland and Acacia. It is furthermore interesting (and somewhat counter-intuitive) to note the changes in atmospheric states associated with high/low line loadings between Acacia and Stikland which likely speaks to the relative positions of these lines within the network topology. Finally, the restrictiveness of using best- and worst-case fictitious planning scenarios in networks with distributed generation is evident for both Stikland and Acacia, based on its (comparatively) broad loading parameters. These parameters may in turn result in the underutilization of network capability, especially in networks with high levels of wind power.

Fig. 7 shows the PPF active power output for two network components – a series reactor and a series capacitor. Similarly to Fig. 6 it is evident from Fig. 7 that the proposed methodology is able to decrease simulated network uncertainty associated with the base state and is able to make sharp differentiations in anticipated active power as a function of the atmospheric state. It is furthermore similarly evident that the best- and worst-case scenarios result in conservative anticipated line loadings which may not be well suited to networks with high levels of wind power penetration. For the Muldersvlei series reactor, the ridging of the Indian- and Atlantic Ocean High Pressure Systems (SOMs (1,1) and (5,1)) are associated with a high probability of high active power (approx. 45 MW) and strong cyclonic circulation (SOM (6,6)) is associated with a high probability of relatively low active power (approx. 25 MW). For the Kronos series capacitor, SOM (1,1) is associated with a high probability of relatively low active power (approx. 210–230 MW) and SOM (5,1) is associated with a high probability of relatively high active power (approx. 360–370 MW). SOM node (1,6) is unable to significantly distinguish itself from the base state. It may thereby be summarized that the proposed methodology is able to define operating scenarios that reduce network uncertainty by increasing the relative probabilities associated with line loadings as a function of atmospheric state occurrence.

5. Conclusions

This paper has set out to formulate a methodology for defining operating scenarios for PPF, based on large-scale atmospheric circulation for networks with significant levels of wind power. The primary underlying hypothesis of this paper was that the probabilistic spatial–temporal information needed to select appropriate operating scenarios to reflect future system states in planning and operational PPFs, may be obtained through the clustering of historic wind power time-series for multiple generators as a

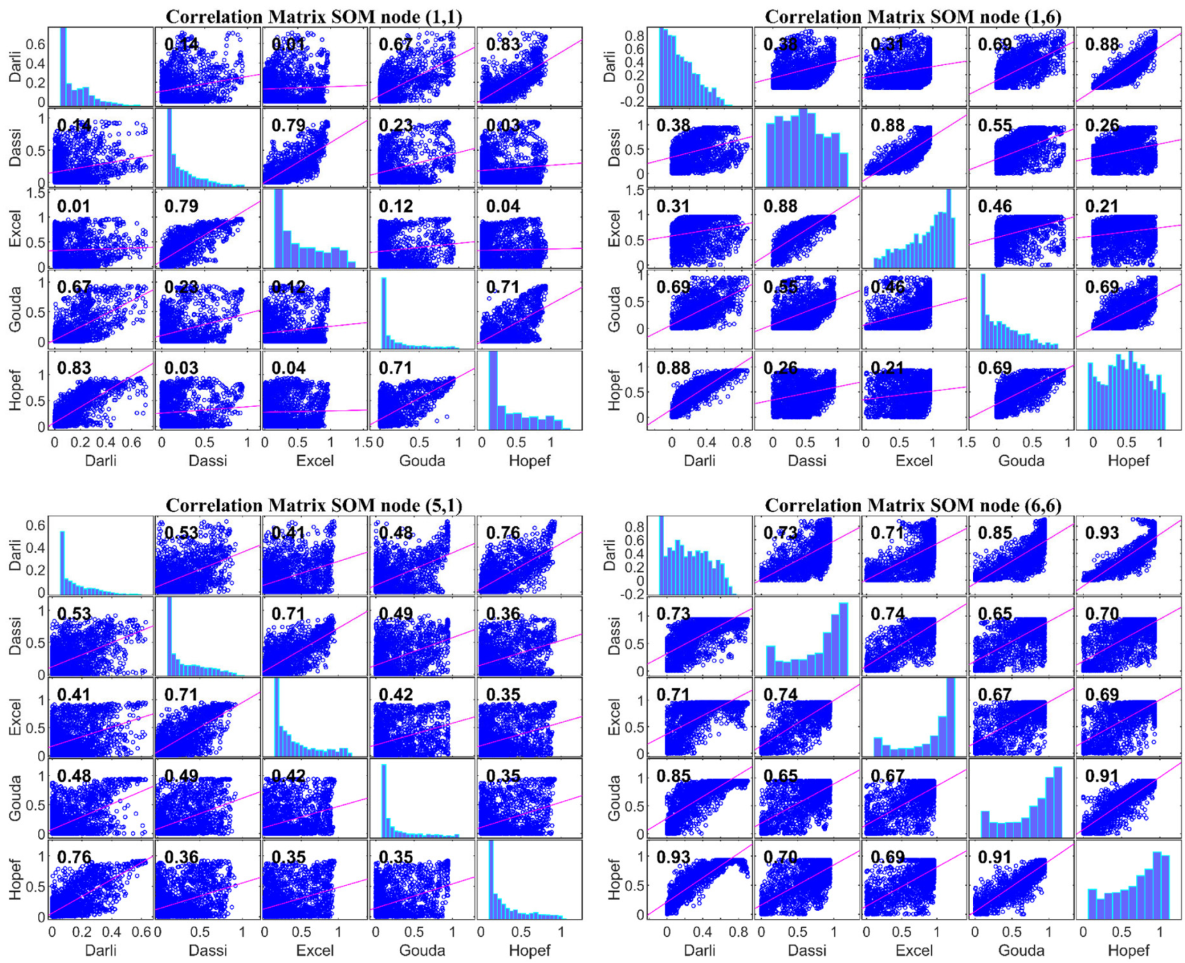

Fig. 5. Correlation Matrices for the four sampled SOM nodes.

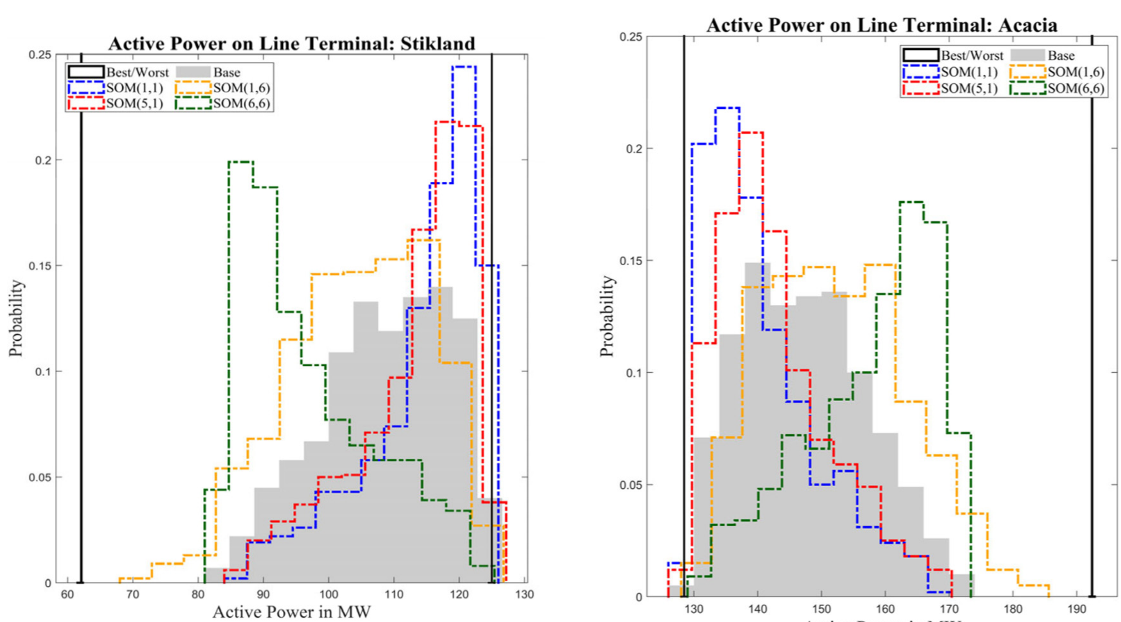

Fig. 6. Active Power on sampled lines within the study area.

function of concurrent classified atmospheric states. Accordingly, for operational PPF, dynamic NWP point-forecasts can be used by network operators to infer the atmospheric state at the forecast time of interest, and thereby the appropriate operating scenario can be selected for PPF to resolve a specific future system state. For planning PPF, frequently occurring SOM nodes, and SOM

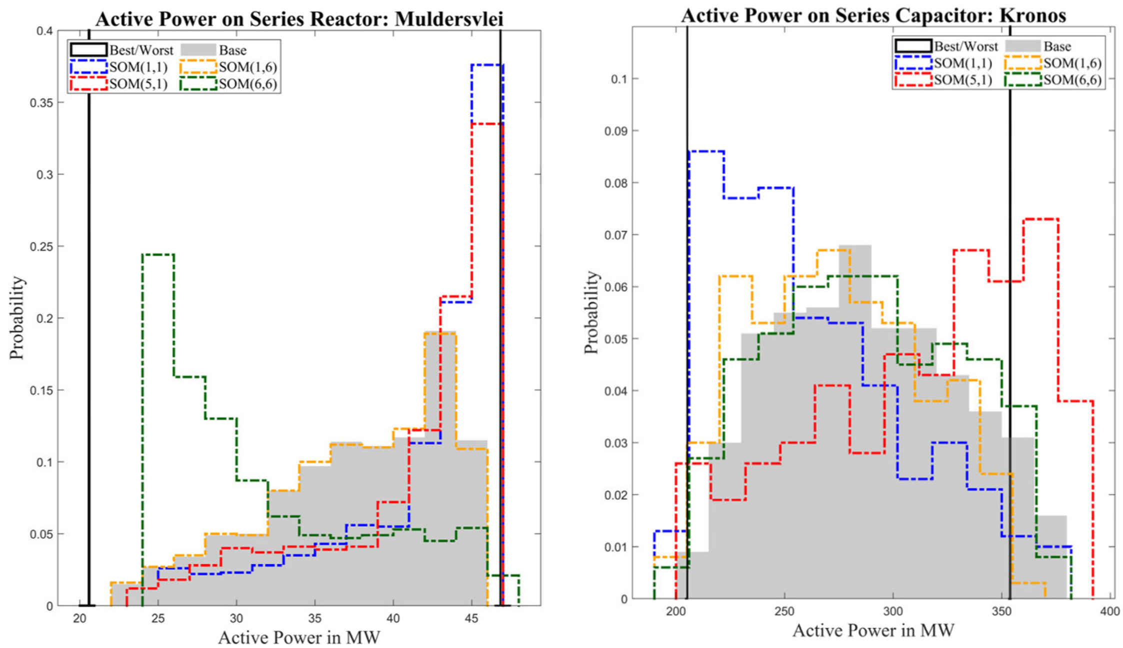

Fig. 7. Active Power on sampled network reactor and capacitor components within the study area.

nodes that represent operating states that the network may find difficult to accommodate, may be used to derive the operating scenarios to include within a network planning and expansion analysis.

To demonstrate its potential benefits, the proposed methodology was applied in a case study considering a portion of the South African transmission network with 11 wind generators. Although one case study is not sufficient to gain a generalized understanding of the relation between certain network states and atmospheric circulation, it is sufficient to demonstrate the contribution of the proposed methodology in reducing simulation uncertainty and representing system states in terms of atmospheric circulation. The following conclusions can be made:

- Sharp distinctions in the probabilistic properties of individual wind power generators and interdependency structures between generators can be elucidated when clustering historic wind power time-series as a function of a set of classified atmospheric states.
- The copula functions derived from the proposed clustering methodology can effectively be applied as a set of operating scenarios within an operational or planning PPF analysis which results in a reduction in simulation uncertainty when compared to an unclustered base case. Such scenarios are furthermore less restrictive than scenarios defined on a highly conservative worst-case planning principle.

The most significant contribution made by this paper is in illustrating how operating scenarios that accurately represent the impact of wind energy generators within a PPF analysis (which includes the probabilistic density functions of generators, and correlations between generators) can be derived from clustering historic generation data as a function of a set of classified atmospheric states. Through its synthesis of atmospheric science and power systems engineering, this paper is interdisciplinary in its approach and its contributions. Future research will be focused on expanding the proposed methodology to other network parameters — notably load demand and solar PV.

CRediT authorship contribution statement

Amaris Dalton: Conceptualization, Methodology, Formal analysis, Writing - original draft. **Bernard Bekker:** Supervision, Writing - review & editing. **Matti Juhani Koivisto:** Software, Resources, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge: Eskom for their willingness to help with this research, specifically in entrusting the authors with their transmission network PowerFactory case files; the Wind Energy Department at the DTU for the use of the CorWind model; and the Centre for Renewable and Sustainable Energy Studies for the financial support provided. Matti Koivisto acknowledges support from the PSfuture (La Cour Fellowship, DTU Wind Energy) project.

References

Al-yahyai, Sultan, Charabi, Yassine, Gastli, Adel, 2010. Review of the use of numerical weather prediction (NWP) models for wind energy assessment. Renew. Sustain. Energy Rev. 14 (9), 3192–3198. http://dx.doi.org/10.1016/j.rser.2010.07.001.

Bent, Russell, Toole, G. Loren, Berscheid, Alan, 2012. Transmission network expansion planning with complex power flow models. IEEE Trans. Power Syst. 27 (2), 904–912.

Borkowska, Barbara, 1974. Probabilistic load flow. IEEE Trans. Power Appar. Syst. PAS-93 (3), 752–759.

Botterud, Audun, Zhou, Zhi, Wang, Jianhui, Bessa, Ricardo J., Keko, Hrvoje, Mendes, Joana, Sumaili, Jean, Miranda, Vladimiro, 2011. Use of Wind Power Forecasting on System Operations. Argonne National Laborotory Report, AN.

BP, 2020. Bp statistical review of world energy 2020. www.bp.com/statisticalreview.

Chen, P., Chen, Z., Bak-Jensen, B., 2008. Probabilistic load flow: A review. In: 2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (April). p. 1586–1591.

Copernicus Climate Change Service, 2017. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. https://cds.climate.copernicus.eu/cdsapp#!/home.

Dalton, Amaris, Bekker, Bernard, Coenrad Kruger, Andries, 2019. Wind power variability during the passage of cold fronts across South Africa. J. Energy South. Afr. 30 (3), 52–67.

Dalton, Amaris, Bekker, Bernard, Juhani Koivisto, Matti, 2020a. Atmospheric circulation archetypes as clustering criteria for wind power inputs into probabilistic power flow analysis. In: 2020 International Conference on Probabilistic Methods Applied to Power Systems. PMAPS. p. 1–6.

Dalton, Amaris, Bekker, Bernard, Juhani Koivisto, Matti, 2020b. Simulation and detection of wind power ramps and identification of their causative atmospheric circulation patterns. Electr. Power Syst. Res. (September), 106936. http://dx.doi.org/10.1016/j.epsr.2020.106936.

DIgSILENT GmbH, 2019. DIgSILENT PowerFactory user manuel. Gomaringen, Germany.

Fitiwi, Desta Z., De Cuadra, F., Olmos, L., Rivier, M., 2015. A new approach of clustering operational states for power network expansion planning problems dealing with RES (renewable energy source) generation operational variability and uncertainty. Energy 90, 1360–1376. http://dx.doi.org/10.1016/j.energy.2015.06.078.

Grams, Christian M., Beerli, Remo, Pfenninger, Stefan, Staffell, Iain, Wernli, Heini, 2018. Europe PMC funders group balancing europe's wind power output through spatial deployment informed by weather regimes. Nature Clim. Change 7 (8), 557–562.

Haghi, H. Valizadeh, Tavakoli Bina, M., Golkar, M.A., 2010. Using copulas for analysis of large datasets in renewable distributed generation: PV and wind power integration in Iran. Renew. Energy 35 (9), 1991–2000. http://dx.doi.org/10.1016/j.renene.2010.01.031.

Jiang, Xichen, Christine Chen, Yu, Domínguez-Garćia, Alejandro D., 2013. A set-theoretic framework to assess the impact of variable generation on the power flow. IEEE Trans. Power Syst. 28 (2), 855–867.

Kiviluoma, Juha, Holttinen, Hannele, Weir, David, Scharff, Richard, Söder, Lennart, Menemenlis, Nickie, Cutululis, Nicolaos A., Danti Lopez, Irene, Lannoye, Eamonn, 2016. Variability in large-scale wind power generation. Wind Energy 19 (9), 1649–1665.

Kohonen, Teuvo, Honkela, Timo, 2007. Kohonen network. Scholarpedia 2 (1).

Kohonen, Teuvo, Kaski, Samuel, Somervuo, Panu, Lagus, Krista, Oja, Merja, 1990. Self-organizing map. In: Proceedings of the IEEE.

Koivisto, Matti, Das, Kaushik, Guo, Feng, Sørensen, Poul, Nuño, Edgar, Cutululis, Nicolaos, Maule, Petr, 2019. Using time series simulation tools for assessing the effects of variable renewable energy generation on power and energy systems. Wiley Interdiscip. Rev.: Energy Environ. 8 (3), 1–15.

Koivisto, Matti, Margrét Jónsdóttir, Guðrún, Sørensen, Poul, Plakas, Konstantinos, Cutululis, Nicolaos, 2020. Combination of meteorological reanalysis data and stochastic simulation for modelling wind generation variability. Renew. Energy 159, 991–999. http://dx.doi.org/10.1016/j.renene.2020.06.033.

Koivisto, Matti, Seppänen, Janne, Mellin, Ilkka, Ekström, Jussi, Millar, John, Mammarella, Ivan, Komppula, Mika, Lehtonen, Matti, 2016. Electrical power and energy systems wind speed modeling using a vector autoregressive process with a time-dependent intercept term. Int. J. Electr. Power Energy Syst. 77, 91–99. http://dx.doi.org/10.1016/j.ijepes.2015.11.027.

Kroposki, Benjamin, Johnson, Brian, Zhang, Yingchen, Gevorgian, Vahan, Denholm, Paul, Hodge, Bri Mathias, Hannegan, Bryan, 2017. Achieving a 100% renewable grid: Operating electric power systems with extremely high levels of variable renewable energy. IEEE Power Energy Mag. 15 (2), 61–73.

Kruger, A.G., Goliger, A.M, Retief, J.V., Sekele, S., 2010. Strong wind climatic zones in South Africa. Wind Struct. Int. J. 13 (1), 37–55.

Larsén, Xiaoli Guo, Ott, Søren, Badger, Jake, Hahmann, Andrea N., Mann, Jakob, 2012. Recipes for correcting the impact of effective mesoscale resolution on the estimation of extreme winds. J. Appl. Meteorol. Climatol. 51 (3), 521–533.

Leite da Silva, A.M., Ribeiro, S.M.P., Arienti, V.L., Allan, R.N., Do Coutto Filho, M.B., 1990. Probabilistic load flow techniques applied to power system expansion planning. IEEE Trans. Power Syst. 5 (4), 1047–1053.

Lennard, Christopher, Hegerl, Gabriele, 2014. Relating changes in synoptic circulation to the surface rainfall response using self-organising maps. Clim. Dynam. 44 (3–4), 861–879.

Leuthold, Florian, Jeske, Till, Weigt, Hannes, von Hirschhausen, Christian, 2008. When the wind blows over europe a simulation analysis and the impact of grid extensions. Electricity Markets Working Papers WP-EM-31 Munich Personal RePEc Archive.

Liu, Yonggang, Weisberg, Robert H., 2011. A Review of Self-Organizing Map Applications in Meteorology and Oceanography. Marine Science Faculty Publications, (419).

Min, Liang, Lee, Stephen T., Zhang, Pei, Rose, Virgil, Cole, James, 2008. Short-term probabilistic transmission congestion forecasting. In: 3rd International Conference on Deregulation and Restructuring and Power Technologies. DRPT 2008 (April). pp. 764–770.

Monforti, F., Huld, T., Bódis, K., Vitali, L., Isidoro, M.D., Lacal-arántegui, R., 2014. Assessing complementarity of wind and solar resources for energy production in Italy. Monte Carlo Approach. Renew. Energy 63, 576–586. http://dx.doi.org/10.1016/j.renene.2013.10.028.

Norgaard, P., Holttinen, H., Nørgaard, P., Holttinen, H., 2004. A multi-turbine power curve approach. In: Nordic Wind Power Conference (March). p. 5, http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Multi-Turbine+Power+Curve+Approach#0.

Openshaw, Stan, 1994. Neural nets: Applications in geography. In: Hewitson, B.C., Crane, R.G. (Eds.), Neural Nets: Applications in Geography. Kluwer, Dordrecht, p. 194.

Raglend, I. Jacob, Padhy, Narayana Prasad, 2006. Solutions to practical unit commitment problems with operational, power flow and environmental constraints. In: 2006 IEEE Power Engineering Society General Meeting. PES 247667. p. 8.

Repo, Sami, Laaksonen, Hannu, 2005. Statistical short-term network planning of distribution system and distributed generation. In: 15th Power Systems Computation Conference (August). pp. 22–26.

Reusch, David B., Alley, Richard B., Hewitson, Bruce C., 2005. Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data. Polar Geogr. 29 (3), 188–212.

Sheridan, Scott C., Lee, Cameron C., 2011. The self-organizing map in synoptic climatological research. Prog. Phys. Geogr. 35, 109–119.

Sklar, Abe, 1973. Kybernetika terms of use: Random variables, joint distribution functions, and copulas*. Kybernetika 9 (6).

Soman, Saurabh S., Zareipour, Hamidreza, Malik, Om, Mandal, Paras, 2010. A review of wind power and wind speed forecasting methods with different time horizons. In: North American Power Symposium 2010. NAPS 2010. pp. 1–8.

Sørensen, Poul, Cutululis, Nicolaos Antonio, Vigueras-Rodríguez, Antonio, Madsen, Henrik, Pinson, Pierre, Jensen, Leo E., Hjerrild, Jesper, Donovan, Martin, 2008. Modelling of power fluctuations from large offshore wind farms. Wind Energy 11 (1), 29–43.

Sørensen, Poul, Hansen, Anca D., André, Pedro, Rosas, Carvalho, 2002. Wind models for simulation of power fluctuations from wind farms. J. Wind Eng. Ind. Aerodyn. 90 (12–15), 1381–1402.

Sørensen, Poul, Litong-palima, Marisciel, Hahman, Andrea N., Heunis, Schalk, Ntusi, Marathon, Hansen, Jens Carsten, 2018. Wind power variability and power system reserves in South Africa. J. Energy South. Afr. 29 (1), 59–71.

Stott, Brian, 1974. Review of load-flow calculation methods. Proc. IEEE 62 (7), 916–929.

Tang, Junjie, Ni, Fei, Ponci, Ferdinanda, Monti, Antonello, 2016. Dimension-adaptive sparse grid interpolation for uncertainty quantification in modern power systems: Probabilistic power flow. IEEE Trans. Power Syst. 31 (2), 907–919.

Tyson, Peter Daughtrey, Preston-Whyte, Robert A., 2000. Weather and Climate of Southern Africa. Oxford.

Ueckerdt, F., Brecha, R., Luderer, G., 2015. Analyzing major challenges of wind and solar variability in power systems. Renew. Energy 81 (2015), 1–10.

Usaola, J., 2009. Probabilistic load flow in systems with wind generation. IET Gener. Transm. Distrib. 3 (12), 1031–1041.

Van Hertem, D., Verboomen, J., Purchala, Konrad, Meeus, Leonardo, Belmans, Ronnie, Kling, W.L, 2005. Usefulness of DC power flow for active power flow analysis with flow controlling devices. In: 2005 IEEE Power Engineering Society General Meeting, Vol. 1. pp. 454–459.

Zhang, Yao, Wang, Jianxue, Wang, Xifan, 2014. Review on probabilistic forecasting of wind power generation. Renew. Sustain. Energy Rev. 32 (2014), 255–270.